

## DARPAN BERI

Ontario, Canada | (249)-535-5398 | [darpanberi@trentu.ca](mailto:darpanberi@trentu.ca) | [LinkedIn.com/in/darpanberi/](https://www.linkedin.com/in/darpanberi/) | [Github.com/DarpanBeri](https://github.com/DarpanBeri)

### EDUCATION

**Master of Science, Applied Modelling and Quantitative Methods - Big Data Analytics** | GPA: 4.0/4.0

**Capstone:** When Machines Evaluate Machines: A Critical Assessment of AI Evaluators in the Context of Optimizing LLaMa Based Retrieval Augmented Generation Pipelines. | [Academic Paper Link](#) | [Poster Link](#)  
Trent University, Peterborough, Ontario, Canada | January, 2024 - May, 2025

**Bachelor of Science, Computer Science** | GPA 3.03/4.0

San Diego State University, San Diego, California, USA | August 2016 - December, 2020

### PROFESSIONAL EXPERIENCE

**Software Engineer**, Gemini Solutions Pvt Ltd, Austin, Texas, USA | December 2021 - June 2023

- Consulted at Pacific Investment Management Company, LLC (PIMCO) to centralize critical data pipelines, resulting in a 10% reduction in big data SQL query usage and 10-fold decrease in ETL errors.
- Developed and automated data pipelines, enabling smooth transfer of big data from kdb+ to Amazon S3, while enhancing performance monitoring, logging, and overall pipeline efficiency.
- Implemented automated validation checks for data transferred to S3 and set up Slack alerts for specific failure cases, enabling prompt team response.

**Data Scientist**, San Diego Association of Governments (SANDAG), San Diego, California, USA | August 2021 - December 2021

- Developed an automated ETL pipeline for monthly infobit calculations, quality checks, and visualization. Drastically improved infobit release time from 30-45 days to just 15 minutes, saving over 40 man-hours. The project was publicly showcased and garnered media attention([KPBS](#), [The San Diego Union-Tribune](#), and [NBC San Diego](#)).

**Data Science and Analytics Intern**, San Diego Association of Governments (SANDAG), San Diego, California, USA | January 2021 - July 2021

- Recognized for surpassing goals and contributing significantly to data initiatives, managing diverse data sources for Chief Economist presentations to the regional interest groups.
- Developed a comprehensive multi-dimensional analysis of San Diego County, leveraging NAICS codes and population segmentation for regional planning.
- Conducted a Census Pulse Survey analysis, producing a report on the economic impact of COVID-19, aiding officials in data-driven decision-making.

### PROJECTS AND LEADERSHIP

**[CiteSmart](#) (HackTrent, 2024):** Developed an AI-powered legal assistant using RAG and LLaMa 3.2 to retrieve and cite answers from the *Canadian Environmental Protection Act*, with fallback to DuckDuckGo API for out-of-scope queries.

**Explainable AI (xAI) Clinical Decision Support System (2024):** Prototyped an explainable machine learning model which predicted hospital admission based on current and previous medical history.

**Founder, San Diego Association of Governments (SANDAG)'s Tea Club (2021)**

**Nominee, San Diego Association of Governments (SANDAG)'s Diversity, Equity, and Inclusion Board (2021)**

**President, San Diego State University's Artificial Intelligence Club (2019)**

### TECHNICAL SKILLS

**Programming Languages:** Python, SQL, R, C++, Java, C, JavaScript

**AI / ML / Data Science:** Scikit-learn, PyTorch, Keras, TensorFlow, Hugging Face Transformers, Haystack, NLTK, OpenCV, Pandas, NumPy, SciPy, Seaborn, Matplotlib,

XGBoost, LLM Prompt Engineering, LLM Tuning, RAG, Statistical Modeling, Hypothesis Testing, Sentence Transformers  
**Cloud / DevOps / Infrastructure / Databases:** AWS S3, Docker, Kubernetes, Git, Linux, Jira,

PagerDuty, Grafana, Microsoft SQL Server Management Studio (SSMS), KDB+, API Integration  
**Data Visualization/BI:** Power BI, ArcGIS, Tableau  
**Collaboration Tools:** Slack